

OPEN

# Advanced identification of global bioactivity hotspots via screening of the metabolic fingerprint of entire ecosystems

Constanze Mueller<sup>1,8\*</sup>, Stephan Kremb<sup>2,5,8</sup>, Michael Gonsior<sup>3</sup>, Ruth Brack-Werner<sup>4</sup>, Christian R. Voolstra<sup>5,6</sup> & Philippe Schmitt-Kopplin<sup>1,3,7\*</sup>

Natural products (NP) are a valuable drug resource. However, NP-inspired drug leads are declining, among other reasons due to high re-discovery rates. We developed a conceptual framework using the metabolic fingerprint of entire ecosystems (MeE) to facilitate the discovery of global bioactivity hotspots. We assessed the MeE of 305 sites of diverse aquatic ecosystems, worldwide. All samples were tested for *antiviral* effects against the human immunodeficiency virus (HIV), followed by a comprehensive screening for cell-modulatory activity by High-Content Screening (HCS). We discovered a very strong HIV-1 inhibition mainly in samples taken from fjords with a strong terrestrial input. Multivariate data integration demonstrated an association of a set of polyphenols with specific biological alterations (endoplasmic reticulum, lysosomes, and NFκB) caused by these samples. Moreover, we found strong HIV-1 inhibition in one unrelated oceanic sample closely matching to HIV-1-inhibitory drugs on a cytological and a chemical level. Taken together, we demonstrate that even without physical purification, a sophisticated strategy of differential filtering, correlation analysis, and multivariate statistics can be employed to guide chemical analysis, to improve de-replication, and to identify ecosystems with promising characteristics as sources for NP discovery.

Natural products (NP) are an attractive starting point for drug screening and discovery due to their high structural diversity and complexity<sup>1–3</sup>. Approximately 60% of all currently registered drugs originate from natural sources, which is particularly true for anti-infectives and cytostatic drugs<sup>1–4</sup>. However, it is becoming increasingly clear that the discovery rate of novel structures from traditional sources of NP is declining<sup>5</sup>. In this context, new original sources of NP can help in providing novel molecular entities for the development of lead compounds and to avoid increasing redundancy and rediscovery rates<sup>6,7</sup>.

In this study, we explored the potential of complex mixtures extracted from different aquatic ecosystems worldwide as novel sources for bioactive compounds. We focused our efforts on complex mixtures collected in aquatic ecosystems as molecules are being leached out of surrounding ecosystems and are being transported into water reservoirs by boundless water cycles (Fig. S1a). Considering the fact that living biota releases signaling or defense molecules into their adjoining environments<sup>8,9</sup>, it is only reasonable that traces of such bioactive molecules can be found in these samples. In general, the complex extracts used in this study contain intact molecules exuded, excreted, leached and otherwise released from any living and decaying biota within the ecosystem<sup>10</sup>. Their chemical richness is further increased by chemical and biochemical transformation and alteration

<sup>1</sup>Research Unit Analytical BioGeoChemistry, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, D-85764, Germany. <sup>2</sup>NYUAD Center for Genomics and Systems Biology, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates. <sup>3</sup>University of Maryland Center of Environmental Science, Chesapeake Biological Laboratory, Salomons, MD, 20688, USA. <sup>4</sup>Institute of Virology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, D-85764, Germany. <sup>5</sup>Red Sea Research Center, Division of Biological and Environmental Science and Engineering (BESE), King Abdullah University of Science and Technology (KAUST), Thuwal, 23955-6900, Saudi Arabia. <sup>6</sup>Department of Biology, University of Konstanz, Konstanz, 78457, Germany. <sup>7</sup>Chair of Analytical Food Chemistry, Technische Universität München, Freising-Weihenstephan, D-85354, Germany. <sup>8</sup>These authors contributed equally: Constanze Mueller and Stephan Kremb. \*email: [constanze.mueller@helmholtz-muenchen.de](mailto:constanze.mueller@helmholtz-muenchen.de); [schmitt-kopplin@helmholtz-muenchen.de](mailto:schmitt-kopplin@helmholtz-muenchen.de)

processes<sup>10</sup>. In consequence and by extension of the definition of the metabolome (entirety of small molecules (<1 kDa) within an organism), these complex extracts can be considered as the metabolome of the entire ecosystem (MeE). Importantly, the sampling procedure of the MeE does not involve ecological impact. Hence, it can easily be applied as pilot screening for drug discovery and creation of a global natural product catalogue.

Due to the high complexity of the MeE (several thousand molecules per sample), in-depth chemical profiling requires ultra-high resolving analytical tools with a very high sensitivity and dynamic range<sup>11,12</sup>. In terms of mass spectrometry, such properties are currently only provided by Fourier transform ion cyclotron resonance mass spectrometers (FT-ICR-MS). Ultra-high resolution (UHR) thereby enables to distinguish between even very closely located  $m/z$  (mass/charge) signals. The extraordinary dynamic range poses the advantage of detection of compounds of very different concentrations and ionization efficiencies. Such analytical performance allows a *non-targeted (unbiased) chemical analysis* led by statistical modeling.

In terms of bioactivity screening, most conventional studies on effects caused by small molecules and NP have focused either on single molecular targets or general toxicity (*targeted biological analysis*)<sup>13</sup>. Unfortunately, the full bioactivity potential might not be detectable using these approaches. Imaging-based High-Content Screening (HCS) using a set of fluorescence dyes or antibodies that target several cellular structures has recently emerged as a promising tool for more holistic screenings<sup>14–19</sup>. It provides a powerful strategy for *non-targeted (unbiased) biological profiling* of alterations caused by small molecules and for prediction of compound-related mode of action (MoA)<sup>14,15</sup>. However, combined analyses of the in-depth chemical composition with comprehensive biological activity profiles are not reported in literature. To our best knowledge only Kurita *et al.*<sup>20</sup> and Kremb *et al.*<sup>21</sup> initiated this type of analysis, but with samples of low complexity and using only a limited set of cellular structures. Another strategy for the identification of NP candidates with promise is the screening for *antiviral* activities, such as inhibition of the human immunodeficiency virus type-1 (HIV-1). This has been successfully demonstrated for a broad variety of complex mixtures of natural products, using a robust phenotypic screening assay encompassing the entire HIV replication cycle (EASY-HIT)<sup>22–26</sup>.

We here present a global survey of combined chemical and biological profiling of diverse MeE to pinpoint environments that should serve as promising starting points in future NP discovery studies. We performed in-depth characterization of the chemical composition of each MeE combined with (i) a well-established *cell-based antiviral* assay (EASY HIT anti-HIV-1) and (ii) with a *comprehensive* High-content Screening, which yields insights into the altered cell physiology of treated mammalian cells. For both assays, we employed diverse MS-based informatics approaches, including multivariate statistics and UHR molecular networking, to link the chemical composition with the obtained bioactivity of the sample.

## Results

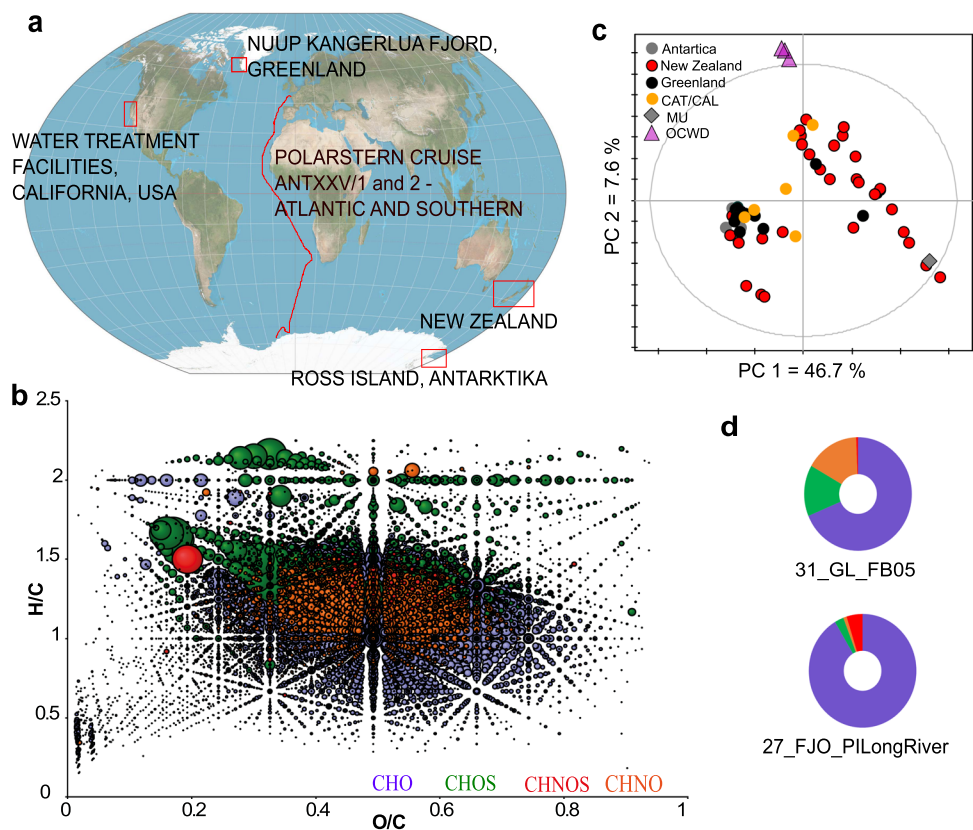
### Chemical characterization of worldwide sampled metabolic fingerprints of entire ecosystems (MeE).

We applied *non-targeted* UHR mass spectrometry to capture the chemical space of 305 MeE samples collected in five continents (Europe, Africa, Australia, North America and Antarctica) at different sites in aquatic ecosystems (Fig. 1a). We included field samples of coastal and marine ecosystems, as well as along vertical and horizontal gradients of several fjords, which link terrestrial and marine ecosystems. The organic material contained in the water samples was concentrated by solid phase extraction (SPE) prior to analysis (Fig. S1b). SPE preparation of the samples furthermore ensured enrichment of molecules in a typical drug hydrophilicity range (logP of approximately  $-0.4$  to  $+5.6$ ). Each single sample yielded a distinct chemical fingerprint, consisting out of several thousand detected  $m/z$  features and their relative intensities (in total >31,000 different  $m/z$  features, Fig. S2). These fingerprints varied between the samples according to the sampling sites and reflect the geo-ecological origin of the samples. The captured chemical space is very broad and the detected  $m/z$  features are distributed in all compound classes (Fig. 1b), but with profound differences according to the sampling site of the MeE (Fig. 1c,d).

### Detection of *antiviral* activity in MeE and characteristics of related ecosystems.

At first, we tested the idea of using the MeE for screening for bioactivity hotspots with an *antiviral* assay. We tested all 305 samples using a cell-based full-replication assay capturing all steps of the HIV-1 life cycle<sup>26</sup>. The fluorescence-based assay uses adherently growing HIV-susceptible cells, with a stable fluorescent reporter gene activated by HIV Tat and Rev. Furthermore, a MTT assay was used to address cell viability upon treatment<sup>26</sup>. Results from virus inhibition assay and MTT tests are given in Table S1. Initially, 88% of the samples showed no effect on HIV infection, 5% ( $n = 16$ ) resulted in a moderate inhibition (20–60% of infected cells compared with untreated cultures), 3% ( $n = 8$ ) in a strong inhibition (10–20% infected cells) and 4% ( $n = 11$ ) exhibited a very potent HIV-1 inhibition (less than 10% infected cells) (Fig. 2a). Only one sample showed a significant reduction of cell viability assessed by the MTT proliferation assay. This sample was collected in a deeper water layer (380 m) of the *Doubtful sound* Fjord (21\_FJO\_DS 380), New Zealand.

We connected the *antiviral* activity with the geo-ecological origin of the samples, and observed that virus inhibition was mainly present in MeE which were either sampled along fjords at specific sampling sites or collected off-shore in unrelated environments. Regarding the first mentioned point, activity was found for instance exclusively in surface waters along the *Doubtful sound* Fjord, while samples taken at the same location but in deeper layers (185 m or 300 m) exhibited no inhibitory effect on HIV-1 replication (Fig. 2b). Similarly, porewater from the organic horizon of forest soils and stream samples from *Dusky sound* also showed potent activity. Deep water samples from within the fjord were inactive, as was observed for other fjords, as well. The *Preservation inlet* surface water, as well as Long River (freshwater) sample discharging in the fjord showed strong anti-HIV activity, while the water collected deeper at 350 m was once again inactive. Due to these observations we reasoned that a terrestrial input is probably the source of the bioactive molecules. We used Pearson correlation to search for single  $m/z$  features or groups of  $m/z$  features which are associated with the observed anti-HIV activity of the

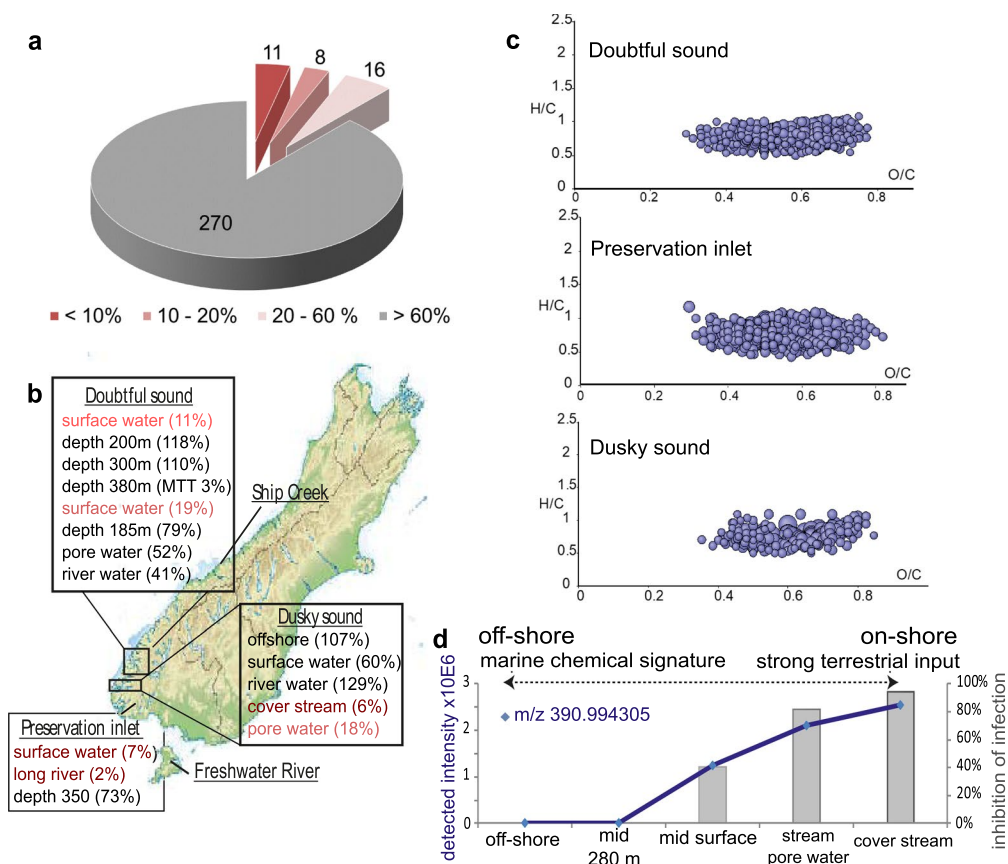


**Figure 1.** Sampling sites and their chemical characterization. An overview of the geographic origin of the analyzed MeE is indicated in the world map (map from Wikipedia, reuse permitted under the Creative Commons Attribution-ShareAlike 3.0 Unported license (CC-BY-SA 3.0, <https://creativecommons.org/licenses/by-sa/3.0/>), created by Strebe, [https://en.wikipedia.org/wiki/World\\_map#/media/File:Winkel\\_triple\\_projection\\_SW.jpg](https://en.wikipedia.org/wiki/World_map#/media/File:Winkel_triple_projection_SW.jpg), modified) (a). All samples were screened for their chemical composition via FT-ICR-MS analysis, which resulted in chemical fingerprints consisting out of several thousand  $m/z$  features per sample and their relative intensities. (b) The combined van Krevelen diagram of all MeE depicts a broad distribution across the chemical space of all elemental compositions (CHO (blue), CHOS (green), CHNO (orange) and CHNOS (red)). Bubble sizes represent the summed detected intensities. (c) PCA score plot based on chemical fingerprints, in which samples are colored according to the geographic location of sampling gives an overview of similarities and differences within the samples' organic content. (d) The chemical space at different sampling sites varied thereby profoundly, exemplarily shown for 31\_GL\_FB05 (taken in Greenland) and 27\_FJO\_PILongRiver (taken in New Zealand). These two samples were chosen as they pose a high *antiviral* activity, which is later described in detail. The ring charts give the quantitative distribution of elemental compositions (color coding according to b).

complex extracts. Correlation coefficient calculations were done *separately* for each sampling site along the depth, and along the course towards the mouth of the fjord using the detected intensity of the single  $m/z$  feature and the observed antiviral potency of the complex extract. We focused on  $m/z$  features with a correlation coefficient  $>0.8$ . For all above described ecosystems a group of putative polyphenols were obtained, exemplarily illustrated for *Dusky sound*, *Doubtful sound* und *Preservation inlet* (Fig. 2c). These molecules show a declining abundance towards the mouth of the fjords (exemplarily illustrated for one  $m/z$  feature in Fig. 2d).

Interestingly, one sample fell in many ways from this above described series of antiviral samples containing a strong terrestrial input: the sample 31\_GL\_FB05. 31\_GL\_FB05 was clearly HIV-active and reduced the virus replication to five percent, but it originates from an off-shore marine ecosystem in Western Greenland, without any terrestrial input. The molecules in this sample that correlate with HIV inhibition contain nitrogen in addition to carbon, hydrogen and oxygen and occupy the van Krevelen region of H/C 1–1.5 and O/C 0.3–0.6 (not shown). Important to note is furthermore, that this sample harbors additionally many molecules, which have been detected in the other Greenland samples, too. Therefore, a narrow clustering of Greenland samples can be recognized in the PCA score plot (Fig. 1c).

Furthermore, we performed a supervised machine learning to capture  $m/z$  features with the highest statistical power for differentiation of anti-HIV active (excluding the marine sample 31\_GL\_FB05 as its chemical fingerprint is very different) and non-active samples. The Partial least square discriminant analysis ((PLS-DA),  $R^2 = 0.914$ ,  $Q^2 = 0.687$ ) delivered a set of 10  $m/z$  features as significantly enriched ( $>3$  fold,  $p < 2.25 \cdot 10^{-5}$  Benjamini/Hochberg corrected for multiple testing) or exclusively present in the most potent *antiviral* active samples. Their



**Figure 2.** Summary of anti-HIV activity of MeE and related  $m/z$  features. All samples were tested for their HIV-1-inhibitory potency. A general overview of obtained results is given in the pie chart (a) (full data can be found in Table S1). We categorized the obtained viral inhibition in 4 groups: less than 10%, 10–20%, 20–60% of infection compared to non-treated infected cells. The number of samples belonging to each group is illustrated. Nine out of eleven samples with very high anti-HIV activity (<10% of infection) originate from fjords in New Zealand (map from Wikipedia, reuse permitted under the Creative Commons Attribution-ShareAlike 3.0 Unported license (CC-BY-SA 3.0, <https://creativecommons.org/licenses/by-sa/3.0/>), [https://en.wikipedia.org/wiki/File:New\\_Zealand\\_relief\\_map.jpg](https://en.wikipedia.org/wiki/File:New_Zealand_relief_map.jpg), modified). (b) We observed a relation of sampling sites within the fjords and HIV inhibition: for all fjords only surface waters and waters with strong terrestrial input show *antiviral* activity (HIV infection is given in % in brackets; *antiviral* samples are highlighted red). In contrast, in waters taken at the same location but from deeper layers the *antiviral* activity was absent. (c) A correlation analysis of the detected intensity of  $m/z$  features and detected HIV inhibition was performed to extract putatively associated  $m/z$  features out of the dataset for each sampling site. For all fjords, a set of CHO containing molecules delivered correlation coefficients >0.8. All of these cover the region of polyphenols in the van Krevelen diagram. (d) Following the correlation analysis between presence of  $m/z$  features and observed HIV inhibition over different depths at one sampling site, we surveyed the presence of these anti-HIV-associated  $m/z$  features along the fjords. This is exemplarily shown for one representative  $m/z$  390.994305. The detected intensity of this  $m/z$  in complex extracts sampled along the fjord is illustrated, as well as the detected HIV inhibition of the total extract. The concentration of the molecule is higher deep inside the fjords, while it is absent at the mouth of the fjords and in related off-shore samples.

molecular composition is given in Table 1. All molecules clustered closely together in the H/C and O/C ratios of 0.5–1.2 and 0.4–0.6 and are uniquely composed of carbon, hydrogen and oxygen. We crosschecked NIAID (National Institute of Allergy and Infectious Diseases ChemDB: Division of AIDS Therapeutics Database), HMDB (Human Metabolome Database)<sup>27,28</sup>, KEGG (Kyoto Encyclopedia of Genes and Genomes)<sup>29</sup> and LIPID MAPS Structure Database (LMSD)<sup>30</sup> for related entries. None of the  $m/z$  features is registered in the NIAID database, the biggest database for infectious diseases. One  $m/z$  feature corresponded to a Quercetin derivative (Quercetin 3-(2'-galloyl- $\alpha$ -L-arabinopyranoside)) within an error of 0.07 ppm in LMSD (Table S2). Moreover, we performed sustained off-resonance irradiation collision-induced dissociation (SORI-CID) of the most abundant  $m/z$  features out of these 10 for further structural information. The procedure allows precise isolation of  $m/z$  features and is the method of choice for complex samples. Ions of interest (parent ions) are isolated, radially accelerated and fragmented by collisions with argon atoms inside the ICR cell, which results in informative patterns of product ions. It mainly led to neutral losses of 18, 32, 44 and 76 (Table 2 SI), which belong to typical fragments such as H<sub>2</sub>O, CO<sub>2</sub>, CH<sub>3</sub>OH and a combination of CH<sub>3</sub>OH and CO<sub>2</sub> (Table S3). The loss of H<sub>2</sub>O indicated the existence of hydroxyl functional groups

m/z	Internal ID	ERROR [ppm]	H	C	O	N	S
343.04594	8069	0.017	11	17	8	0	0
527.08312	7109	-0.056	19	25	13	0	0
529.06238	8124	-0.084	17	24	14	0	0
529.09877	7061	-0.037	21	25	13	0	0
543.07803	7821	-0.063	19	25	14	0	0
559.07295	7115	-0.07	19	25	15	0	0
559.10933	2029	-0.062	23	26	14	0	0
571.07295	7826	-0.086	19	26	15	0	0
573.12498	7453	-0.043	25	27	14	0	0
585.0886	8395	-0.067	21	27	15	0	0

**Table 1.** Most important variables in projection. Common m/z features, which were extracted from the PLS-DA analysis as discriminant for HIV-inhibitory samples. Elemental composition [M-H]<sup>-</sup>, which were calculated using formulae calculator<sup>55</sup>, contain exclusively carbon, hydrogen and oxygen.

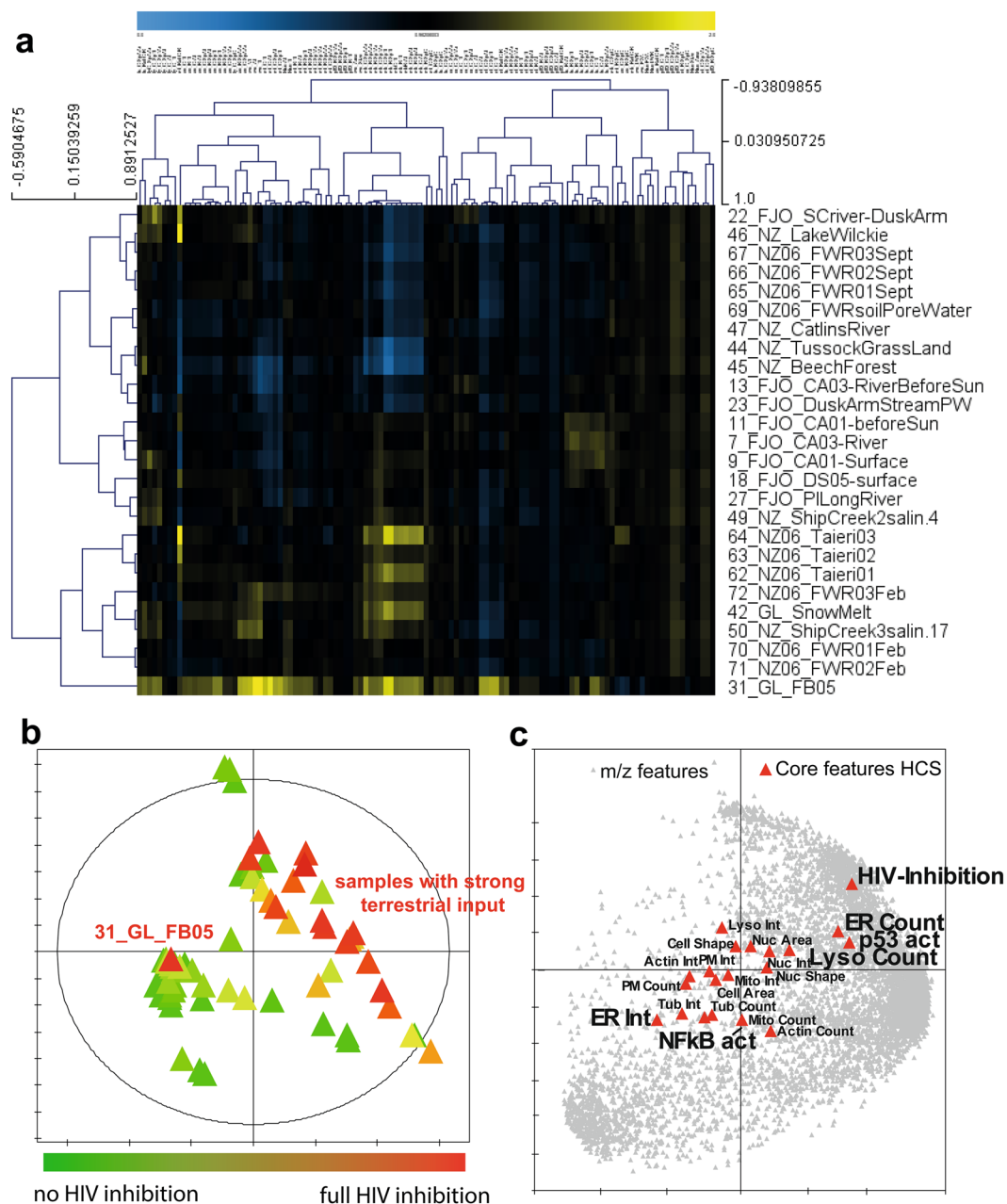
whereas carboxyl groups were most likely present due to the loss of CO<sub>2</sub><sup>31</sup>. Taking these fragmentation patterns together with the localization of the m/z features in the van Krevelen diagram (H/C 0.69–0.84 and O/C 0.52–0.62)<sup>12</sup>, the data strongly suggest the importance of highly oxidized polyphenols.

**In-depth phenotyping of antiviral MeE using HCS.** Following the assessment of antiviral activity of the complex extracts, a comprehensive High-content Screening (HCS) approach was used, which provides an unbiased hypothesis-free biological characterization of alterations caused by the complex extracts on a single-cell level. The HCS platform combines a set of fluorescent dyes targeting 11 cellular structures (nucleus, actin, tubulin, mitochondria, whole cell, endoplasmic reticulum, lysosomes, membranes, NFκB, Caspase 9, p53) resulting in 134 cellular measures and a characteristic cytological profile. This cytological profile covers subsequently a broad spectrum of cell physiology and can serve as a unique fingerprint for sample-induced or drug-induced alterations on mammalian cells<sup>15</sup>. Importantly, drugs that share a MoA deliver comparable cytological profiles, which allows the use of HCS as a prediction tool<sup>14,15</sup>. We screened ecosystems which were found to contain at least one antiviral sample. To cross-reference their cytological profiles to known MoA, we compared cytological profiles obtained from the MeE to a library of 720 single bioactive reference compounds selected from the LOPAC<sup>®</sup>1280 library of pharmacologically active compounds (Sigma Aldrich, international version, <http://www.sigmaaldrich.com/life-science/cell-biology/bioactive-small-molecules/lopac1280-navigator.html>)<sup>15,32,33</sup>. Detected cytological profiles of MeE taken at comparable sample sites within ecosystems are largely consistent (e.g. samples 67, 68, 69, NZ-FW; 62–64, NZ Taieri, Fig. S4). This indicates a common pool of bioactive molecules for these MeE. HIV-inhibitory samples yielded in general four major clusters with multiple sub-clusters by two-dimensional Spearman rank clustering (Fig. 3a). While the first three clusters are mainly a result of differently regulated membrane features, the last cluster shows a high variation from the others over all cytological features. The last cluster consists exclusively out of the sample 31\_GL\_FB05. Therefore, at least two different MoA are inferential.

Non-targeted HCS, targeted anti-HIV screening results and chemical fingerprints were then integrated in an unsupervised statistical model to extract influenced cellular key parameters and m/z features. Importantly, the different types of data were combined through unit variance scaling and mean centering. A principal component analysis (PCA) score plot was generated, in which samples are colored according to their virus inhibition (Fig. 3b). Clustering between MeE correlated to the properties of the sampling site, with a lesser dependence on their geographical location. HIV-inhibitory samples cluster closely within the first quadrant, where samples taken in ecosystems with a strong terrestrial input are located. Only one *antiviral* active sample was plotted outside this cluster with this non-supervised procedure: The open water sample 31\_GL\_FB05 representing a marine ecosystem off-shore in Greenland.

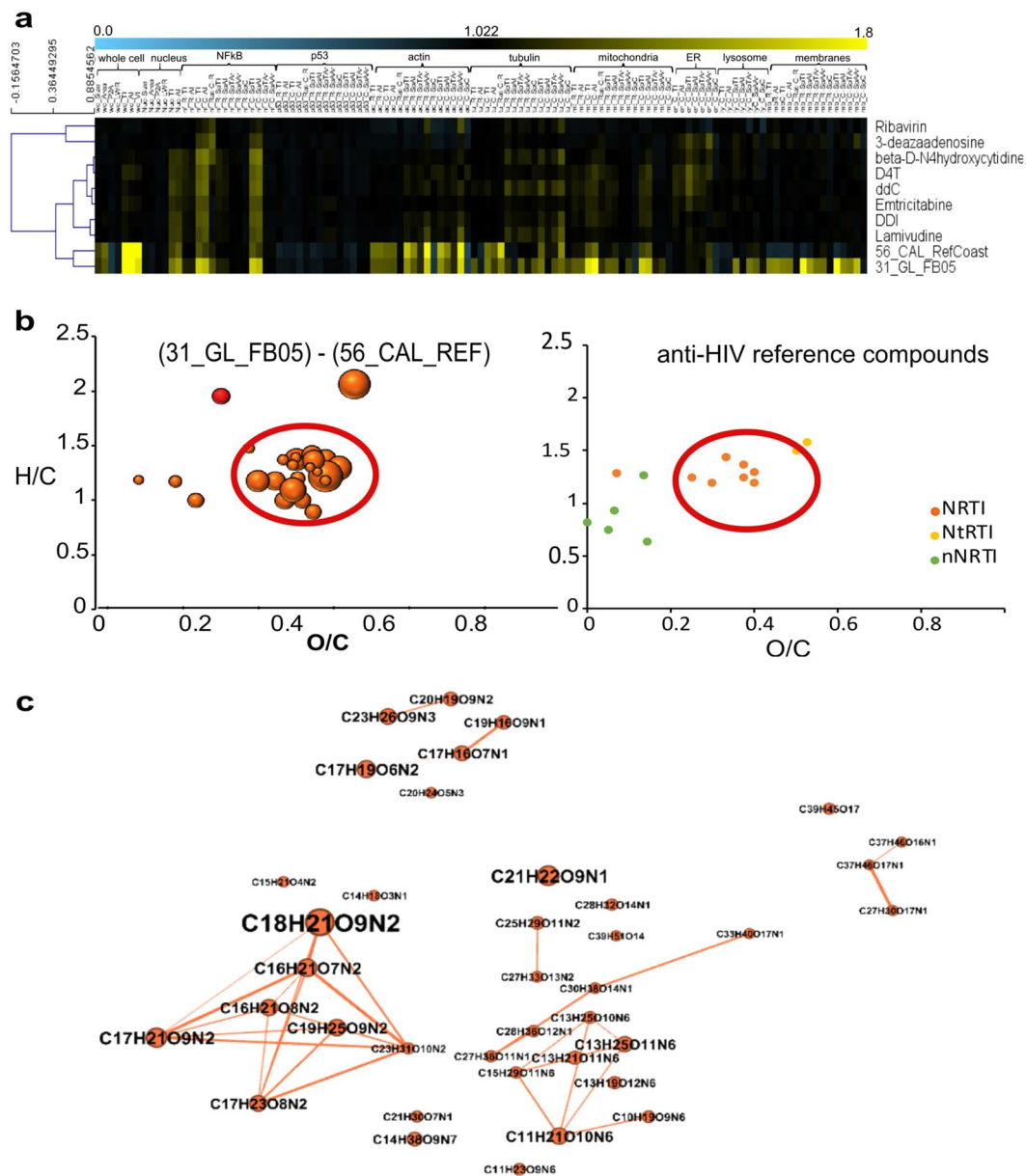
The PCA loading plot (Fig. 3c) demonstrate a small number of cytological parameters, that are modified in cells treated with HIV-inhibitory mixtures. Core features of endoplasmic reticulum (ER), lysosomes, p53 and NFκB (nuclear factor kappa-light-chain-enhancer of activated B cells) are associated with HIV inhibition (Fig. S5a,b). A set of CHO containing compounds, highly likely to be polyphenols (Fig. S5c), followed this correlation for ER-, lysosomal, p53-related features. As an *in-silico* extraction of compounds correlating with NFκB expression over the entire dataset delivered only poor correlation coefficients, we focused here on a subset of samples with a clear effect on NFκB intensity (Fig. S5b). Plant-derived classes of compounds have been positively evaluated as inhibitors of the NFκB pathway<sup>34</sup> and NFκB activation is known to be associated with HIV<sup>34,35</sup>. The correlation analysis between intensities of m/z features and NFκB total intensity over samples, which show a clear effect on NFκB, gave valid correlation coefficients (>0.8). The molecular composition of these compounds strongly suggest their polyphenolic nature (not shown).

The open water sample 31\_GL\_FB05 showed a highly unique cytological profile with broad activities across almost all observations (Fig. 3a). This particular sample also exhibited strong inhibition of HIV-1 replication. In the secondary analysis, we matched cytological profiles obtained from the samples to a library of more than 500 compounds with assigned MoA, including anti-HIV activity. The cytological profile of the potent *antiviral* MeE clustered closely to a set of FDA-approved nucleosidic HIV-1 reverse transcriptase inhibitors (NRTIs) pointing to the presence of molecules with anti-HIV-1 RT (reverse transcriptase) activity (Fig. 4). Interestingly, a second



**Figure 3.** Obtained HCS results and integration with HIV inhibition and chemical fingerprints. Combined data analysis of the obtained HCS cytological profiles, anti-HIV-1 activity, and chemical fingerprints strongly suggests that the two main groups of *antiviral* MeE discovered in this study, cause virus inhibition through different MoAs and different sets of molecules. **(a)** Cluster analysis of all cytological profiles of *antiviral* samples (>50% inhibition) delivered two main clusters. Importantly, the open water sample 31\_GL\_FB05 shows a very different cytological profile compared to the other *antiviral* samples. (Colors indicate positive (yellow) or negative (blue) deviation from the mean of untreated control cells for each cellular feature (control = 1). Spearman rank correlation was used as a distance metric). **(b)** PCA score plot based on chemical fingerprints, in which samples are colored according to their HIV inhibition, show a close clustering of most *antiviral* samples in the first quadrant. These samples contain a strong terrestrial input and a related chemical fingerprint. Only one *antiviral* sample, 31\_GL\_FB05, does not belong in this cluster, which indicates that this sample contains a different set of molecules. **(c)** The PCA loading plot of combined data (HCS cytological profiles, HIV-1 inhibition, UHR mass spectrometric fingerprints) shows an overview of cellular core features, which are correlated/anti-correlated with the HIV inhibition and with a subset of m/z features. For easier readability the HCS parameters were limited to a set of core markers, which contains most of the information.

sample (56\_CAL\_RefCoast) matched closely with the same set of RT-inhibitory reference compounds (Fig. 4a). However, this sample did not exhibit *antiviral* activity in our assay. While both samples show high correlation across profiles, we found strong differences particularly in the membrane marker region, with the HIV-1 active



**Figure 4.** Detection of NRTI-like signatures in a marine *antiviral* MeE. The marine sample 31\_GL\_FB05 shows a close clustering to HIV reference compounds (NRTI). *M/z* features, which are uniquely present in this sample, cover the same chemical space as these NRTIs and show related elemental composition. **(a)** Secondary analysis integrating reference compounds and cytological profiles obtained from MeE delivers a close clustering of reference nucleosidic inhibitors of the HIV-1 RT with two MeE samples: the anti-HIV active sample 31\_GL\_FB05 and 56\_CAL\_ref coast. These two samples differ particularly in the membrane marker region, with the HIV-1 active sample showing strong positive deviation. Colors indicate positive (yellow) or negative (blue) deviation from the mean of untreated control cells for each cellular feature (control = 1). Spearman rank correlation was used as a distance metric. **(b)** *M/z* features were filtered for their presence in 31\_GL\_FB05 and absence in 56\_CAL\_RefCoast (which clusters with sample 31\_GL\_FB05 by cytological profiles but does not show HIV-1 inhibition). Filtered *m/z* features cover the same chemical space as NRTIs (CHNO) (orange), CHNOS (red), bubble sizes indicate detected intensity, NRTIs - Nucleoside analog reverse-transcriptase inhibitors, NtRTIs - Nucleotide analog reverse-transcriptase inhibitors, nNRTIs - Non-nucleoside reverse-transcriptase inhibitors. **(c)** Network analysis of these *m/z* features delivered three molecular families given by their elemental compositions. All of the correlated molecules contain nitrogen.

sample showing strong positive deviation. We compared the chemical fingerprints of these two samples. The anti-HIV biologically active open water sample harbors a unique set of specific nitrogen containing molecules (Fig. 4b), which cover the same space as NRTIs in the van Krevelen diagram. A network analysis of the UHR *m/z* feature revealed a partial chemical relation of these compositions (Fig. 4c).

## Discussion

NP are indispensable to modern medicine<sup>1–3</sup>. Nevertheless, discovery rates of novel chemical entities from traditional sources of NP are decreasing<sup>5</sup>. We therefore propose the use of the MeE in pilot screenings for localization of biodiversity hotspots of bioactive small molecules. With this first proof of principle study we could successfully detect distinct biological activities of this type of samples. We were able to describe characteristics that pinpoint ecosystems with high HIV-inhibitory potential and successfully correlated series of molecules in these specific environments. Although physicochemical isolation of single molecules is theoretically possible, we focus here on a strategy to accelerate the search for bioactive NP hotspot-locations. To our knowledge this is the first instance of combining UHR mass spectrometry data with an *unbiased* HIV full replication assay and a comprehensive *unbiased hypothesis-free* cellular phenotyping approach for such complex mixtures.

We assessed the bioactivity potential of MeE using cell-based screening assays as they mimic relevant *in vivo* conditions and allow for the simultaneous testing of effects on multiple steps in complex biological systems<sup>22,26,36</sup>. At first, an *antiviral* testing was used to validate our working hypothesis. We observed two kinds of ecosystems with an *antiviral* activity. On the one hand, samples collected in fjords with strong terrestrial influence showed potent anti-HIV activities. The Fiordland National Park samples were extracted along several fjords, which are coastal ecosystems that are strongly influenced by freshwater. Extreme levels of annual precipitation and steep mountain ranges rising from sea level result in a rich terrestrial input<sup>37</sup>. Interestingly, New Zealand's forests are largely comprised of endemic species, which is especially true for woody plants and which likely contributes to the unique composition of the organic matter found in the coastal water bodies around New Zealand<sup>38</sup>. Statistical tests highlighted a set of putative polyphenols potentially responsible for the antiviral effect. Additional support of a terrestrial origin of anti-HIV active substances were found in streams sampled around *Beech Forest* and *Ship Creek* and in *Lake Wilkie* (*East coast South Island, NZ*), as well as along the *Freshwater River* draining lowland located on Stewart Island. HIV inhibition was observed to be inversely correlated with increasing salinity<sup>37</sup>, which again demonstrates a mainly terrestrial origin of active substances, rather than a marine source. On the other hand, an unrelated marine sample collected off-shore in Greenland showed a strong virus inhibition. Here, a different set of molecules correlated with HIV inhibition.

We then expanded our study with a comprehensive phenotypic profiling. *Non-targeted* imaging-based HCS yielded for the first time detailed insights into bioactivity profiles of such complex natural material. High-content Screening is still a relatively new technology, especially regarding the combination of several dyes. It can be applied to both, single molecules as well as complex mixtures of molecules<sup>14,15,18,20,39</sup>. HCS provides rich information on multiple levels of cell physiology<sup>40</sup> and is able to predict compound-related MoA by comparing cellular phenotypes characterized by cytological profiles to collections of molecules with known effects on multiple target classes<sup>32,33</sup>. We intended to address two goals with this additional screening step: (i) information regarding the putative MoA and (ii) expansion of our analytical approach from a combination of *targeted antiviral/non-targeted* chemical profiling towards *non-targeted/non-targeted* profiling. For both points we got positive results. The detected cytological profiles followed again a clear geo-ecological trend. On the one hand, we detected largely consistent cytological profiles for intriguing samples with a strong terrestrial input, as it is the case after snow melting or deep inside fjords. For those, the correlation of *antiviral* activity with multifactorial high-resolution phenotypic profiles allowed for the detection of multiple cellular features (like ER, lysosomes, and NFκB pathway) specifically affected. A positive correlation of lysosomal induction (increased lysosomal spot area) with HIV-1 inhibition was apparent, especially in the samples with highest *antiviral* activity. It is likely that increased lysosomal activity eventually leads to the degradation of viral components and thereby diminishing productive viral infection, as inhibition of lysosomal activity might result in an increase of HIV infectivity via the endocytic pathway<sup>41</sup>. We clearly observed polyphenols to be characteristic for samples that induced lysosomal effects. This is in agreement with the previous finding that the protective effects of polyphenols are in part due to a modulation of the lysosomal catabolic process in a concentration-dependent manner<sup>42</sup>. Furthermore, we found a negative correlation of two ER-related features (ER spot area and ER total Intensity, Fig. S5a,b) with *antiviral* activity. The ER plays an important role in the HIV-1 secretory pathway and it is possible that compounds targeting this step in the HIV-1 replication cycle lead to reduced HIV-1 infectivity. Moreover, we found a negative correlation of NFκB total intensity with *antiviral* activity. It is well known that HIV-1 gene expression is stimulated by NFκB binding to LTR elements and that, in turn, reduced levels of NFκB decreases HIV-1 proviral expression<sup>35</sup>. On the other hand, we found a marine sample that showed a strong correlation with cytological profiles of therapeutically used NRTIs. Here, a set of nitrogen containing molecules was differentially filtered and might reflect an algal bloom at the Greenland off-shore site. Notably, the differentially filtered nitrogen containing molecules occupy the same chemical space as NRTIs.

## Conclusions

In the presented proof-of-concept study we assessed functional and chemical aspects of complex mixtures from various ecosystems worldwide on a single cell level and in an omics scale. We illustrate that even without physical isolation of unique compounds<sup>43–47</sup> a sophisticated strategy of differential filtering, correlation analysis and multivariate statistics can be used to guide chemical analysis and sample prioritization. We showed for the first time that natural complex samples simply taken by water collection in ecosystems harbor distinct biological activities and can be used to catalogue NP worldwide. As sampling of these kind of mixtures does not harm the ecosystem we suggest to use them for pilot screening to identify most productive sites for future novel NP efforts. Furthermore, we believe that the promising combination of technologies of *non-targeted* biological and *non-targeted* chemical analysis will strongly accelerate future bioactivity studies.

## Methods

**Sample material.** In this study, we selected complex geochemical mixtures of various nature (marine, terrestrial, man-altered) and with a high diversity in their properties.

The sampling area and optical properties of water samples collected in *Fiordland National Park (NZ)* was previously described in detail<sup>37,48</sup>. Samples were collected at various depths within *Doubtful sound*, *Dusky sound* and *Preservation inlet*. Sampling was undertaken in Austral winter, but the low salinity layer (LSL) was still present which demonstrates the high amount of freshwater transported into the fjords, even under unstable stratification conditions. Photochemical effects were likely to be very minimal and the collected material was mainly subject to microbial transformations. Surface fjord waters started to freeze over at some places and water temperature ranged from 0 °C at the surface to 12 °C at depth in the fjords. Dissolved organic matter (DOM) in New Zealand fjords is derived from pristine low elevation evergreen temperate rain forests and higher elevation evergreen beech forests. The production of organic matter, leached out from soils by high rainfall volumes, is enormous and strongly colors the LSL that sits on top of clear ocean waters<sup>37,48</sup>.

Further samples were collected in the *Catlins region (NZ)* and the *Catlin River (NZ)*. This river drains a low elevation temperate rainforest on the East Coast of the South Island of New Zealand. *Lake Wilkie* is also located in the Catlins and is a coastal highly colored small lake dominated by sphagnum moss species. Samples from the low elevation temperate rainforest were collected on the *West Coast of the South Island (NZ)* in the so-called *Ship Creek*, where freshwater is mixing quickly with seawater in this very small estuary, and on *Stewart Island* along a salinity gradient in the Freshwater River. To be able to constrain higher elevation sources of DOM, we also included samples from a stream draining *Tussock grassland* and from an evergreen beech forest also located on the South Island of New Zealand.

Additionally, samples were collected during the expedition ANTXXV/1 and 2 of R/V *Polarstern* along a *transect in the East Atlantic Ocean and the Atlantic sector of the Southern Ocean*. Samples of surface and deeper waters were filtered and both, the water bodies as well as the filters were tested for their HIV-1-inhibitory potential. Details on the samples are given in Schmitt-Kopplin *et al.*<sup>49</sup> and Ksionzek *et al.*<sup>50</sup>.

A third region of sampling was located in *Antarctica*. Samples were collected from underneath the shelf ice of Ross Island.

Furthermore, sampling was done along the *Nuup Kangerlua Fjord, Greenland*<sup>51</sup>.

A last set of samples is of different nature. We included DOM of secondary treated effluent: (i) from the Orange County Sanitation District after it had passed through the microfiltration process of the advanced water purification facility and indirect water reuse facility at Orange County Water District (OCWD)<sup>51</sup>; (ii) from a deep aquifer raw water that is used as source water to the Mesa Utilities drinking water treatment plant located in Costa Mesa, California; and (iii) from a sewage leakage on Catalina Island, California (CAT/CAL).

**Sample preparation.** Samples were directly prepared at the sampling sites and thereafter stored on ice or at –20 °C until analysis. Sampling procedures are described in previous publications<sup>37,48–51</sup>. All samples were filtered using Whatman GF/F glass fiber filters, acidified to pH 2 with high purity hydrochloric acid and then extracted/enriched by SPE according to previously published protocols<sup>52,53</sup>. SPE was performed using Agilent Bond Elut PPL cartridges containing a functionalized styrene-divinylbenzene polymer. Following the standard operation procedure, the cartridge volume was adapted according to the sample volume and according to the DOM concentration (5 g or 1 g PPL resin) to achieve a maximum concentration of organic material<sup>53</sup>.

**Ultrahigh resolution mass spectrometry and raw data processing.** Ultrahigh-resolution mass spectra were acquired on a Bruker solariX ion cyclotron resonance Fourier transform MS (Bremen, Germany) equipped with a 12 Tesla superconducting magnet and an Apollo II source in negative electrospray ionization mode. Samples were injected with a constant flow rate of 120 µl/h, nebulizer gas pressure of 2.2 bar and drying gas pressure of 4 bar at 200 °C. Accumulations time was 0.4 sec. The applied ESI voltage were 3600 V capillary voltage and –500 V end plate offset. The spectra were acquired using a time transient of 4 MW. MS parameters were optimized to reach a maximum of sensitivity in the m/z range 120–800. Transfer optic parameters were therefore ToF 0.6 msec, frequency 4 MHz and RF amplitude of 175 Vpp. 500 scans were acquired for each sample.

An injection of the reference material from the International Humic Substance Society (IHSS) Suwanne River Fulvic Acid was used for quality control at the beginning of each batch analysis. Spectra were externally calibrated first on clusters of arginine (2.5 mg/l in methanol) and internally calibrated on a fatty acid reference list in negative mode; calibration errors in the relevant m/z range were always below 100 ppb. The spectra were exported from Data Analysis using a Signal/Noise (S/N) minimum of 2 and aligned through an in-house written software using a maximum discrepancy of 1 ppm; the data were stored in a matrix. Missing values were imputed with random variables, reflecting minimal values in the data set. Therefore, a vector was built, which contain the minimum detected intensity for each sample over the entire dataset. Missing values were afterwards imputed by randomly generated integers included in this vector (Excel, 2016, Microsoft). Signal intensities were normalized to the sum of intensities of all detected m/z features per sample<sup>54</sup>.

Elemental compositions of detected m/z feature were calculated using the in-house written formulae calculator considering H, C, N, O and S (maximal error <0.2 ppm)<sup>55</sup>. Elemental compositions were filtered according to the seven golden rules, which consider restrictions for the number of elements, LEWIS and SENIOR chemical rules, isotopic patterns, hydrogen/carbon ratios, elemental ratio of nitrogen, oxygen, phosphor, and sulphur versus carbon, element ratio probabilities and presence of trimethylsilylated compounds<sup>56</sup>. Single charged m/z features with at least one isotopic feature were considered valid. Data from all analyzed samples were visualized in van Krevelen diagrams, in which hydrogen-to-carbon (H/C) atomic ratios (y-axis) against oxygen-to-carbon (O/C) atomic ratios (x-axis) are plotted. Hereby, H/C ratio reflects relative aliphaticity and aromaticity (double bond equivalents – DBE), whereas the O/C ratio relates to oxygenation (oxidation and reduction processes)<sup>12</sup>.

The benefit of such visualization is an easier and faster interpretability of complex mass spectra as areas of certain compound classes can be defined<sup>12</sup>. Even though attention has to be paid, that this interpretation is not a chemical identification. Detected intensities are represented by bubble sizes.

Database search was done in MassTRIX allowing negatively charged masses (M-H)<sup>-</sup> and a maximal ppm error of 1 ppm as only restrictions (<http://masstrix3.helmholtz-muenchen.de/masstrix3/57,58>) and a NIAID (National Institute of Allergy and Infectious Diseases ChemDB: Division of AIDS Therapeutics Database, <http://chemdb.niaid.nih.gov>) survey was performed manually for all important elemental compositions.

**Fragmentation experiments of relevant m/z features.** M/z features which were under the top 100 VIPs and which showed at least a fold change of 3 difference between HIV active and non-active samples were selected for fragmentation experiments (n = 10). SORI-CID (sustained off-resonance irradiation collision-induced dissociation) was performed in negative electrospray mode. Mass spectra consisting of 20 scans were acquired on parent and fragment ions (Table S3). SORI power and isolation power were adapted according to requirements of the m/z feature.

**Cell culture.** HEK293T (for virus stock preparation), LC5-RIC (for EASY HIT assay) and HeLa cells (for HCS) and their culture are described in<sup>15,26</sup>. Briefly, HEK293T cells were seeded into 6-well plates at 10<sup>5</sup> cells in 2 ml of cell very-low-endotoxin (VLE)-RPMI 1640 medium (Biochrom AG, Berlin, Germany) supplemented with 10% fetal bovine serum (Biochrom AG) and 1% antibiotic-antimycotic solution (Gibco) before transfection. LC5-RIC cells contain a stably integrated reporter gene for expression of DsRed upon HIV-1 infection. The cells were cultured in cell culture flasks (T-185 Nunc Solo Flask, Nunc International, Wiesbaden, Germany) and maintained at 37 °C, 5% CO<sub>2</sub> using DMEM (Gibco, Karlsruhe, Germany) culture medium supplemented with GlutaMAX<sup>TM</sup>-I (L-alanyl-L-glutamine), pyruvate and 4.5 g/l glucose (all from Gibco, Karlsruhe, Germany), 10% fetal bovine serum (FBS, Biochrom AG, Berlin, Germany), 1% antibiotic-antimycotic solution (Gibco, Karlsruhe, Germany), 1% sodium pyruvate (Gibco, Karlsruhe, Germany). In addition, 0.74 mg/ml geneticin (G418 sulfate; PAA Laboratories, Pasching, Austria) and 0.125 mg/ml Hygromycin B (PAA Laboratories, Pasching, Austria) were added every second passage, to ensure stable expression of the CD4 receptor and general stability of the reporter construct. The cells were kept for a minimum of 48 h in medium without geneticin and hygromycin B prior to the experiments. HeLa cells (parental HeLa cell line, ATCC<sup>®</sup> CCL-2<sup>TM</sup>) were seeded into 384-well plates at a density of 2,000 cells per well in a volume of 25 µl of Dulbecco's modified Eagle medium (DMEM containing GlutaMAX-1; Life Technologies) containing 10% fetal bovine serum (Life Technologies) and 1% antibiotic-antimycotic solution (Life Technologies) and cultured under 37 °C and 5% CO<sub>2</sub> for 24 h prior treatment with complex mixtures. The cells were used for a maximum of ten passages.

**EASY HIT- HIV full replication assay.** A full description of the assay procedure and virus stock preparation used in this study is detailed in Kremb *et al.*<sup>26</sup>. In brief, 24 h after seeding the cells into microwell plates, the cell culture medium was removed and the sample (dissolved in 100 µl cell culture medium) and 20 µl virus inoculum were added to the cells. Samples were prepared by drying of 10 µl of the original solution and re-dissolving in 330 µl of cell culture medium. 100 µl sample was added per well. Experiments were performed in triplicates. Cells were incubated for 48 h after addition of sample and virus inoculum. Cell culture supernatants were then removed from the treatment plate and fluorescence intensities of each culture measured. To measure effects of treatment on production levels of infectious virus, 20 µl of culture supernatants from treatment plates were transferred to uninfected LC5-RIC cells seeded in 96-well plates, plates incubated for 72 h and fluorescent intensities measured. Fluorescent measurements were performed with a Tecan infinite M200 (Tecan, Crailsheim, Germany) at the monochromator wavelengths of 552 nm for excitation and 596 nm for emission.

Subsequently, viability of LC5-RIC cells was determined via a MTT test (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide) by incubation of cell cultures with 50 µg of MTT solution (Sigma-Aldrich, Taufkirchen, Germany) in 100 µl of cell culture medium for 2 h at 37 °C as described in<sup>26</sup>.

**High-content screening.** High-content Screening of cellular alterations upon MeE-treatment was performed according to a published protocol<sup>15</sup> targeting 11 cellular structures (Table 2) and resulting in unique cytological profiles of 134 features. Cells were treated with 25 µl of samples (re-dissolved in DMEM cell culture medium) in 4 replicates. After 24 h of treatment four different cell-staining protocols were applied (described in detail in Kremb *et al.*<sup>15</sup>). A Cellomics ArrayScan VTI (Thermo Fisher Scientific) platform equipped with a 10x objective (Zeiss Plan Neofluar, NA 0.3) was used. Images were analyzed using the Compartmental Analysis Bio Application (Cellomics, Thermo Fisher Scientific). A minimum of 500 valid objects were analyzed per well. Cell cycle analysis and analysis of cell loss were accomplished by using the Cell Cycle Bio Application (Cellomics, Thermo Fisher Scientific) using at least 2000 valid objects. Raw data from automated image analysis for each cytological feature were related to corresponding values from control wells where the control was set to 1. All cytological features of a given fraction or reference compounds (LOPAC<sup>®</sup>1280 – The Library of Pharmacologically Active Compounds, Sigma-Aldrich, Taufkirchen, Germany) were combined to result in a cytological profile.

**Statistical analysis.** Pearson correlation was used to screen for associated m/z features with the observed anti-HIV-1 activity of the complex extracts (Excel 2016, Microsoft, USA). Therefore, the correlation coefficient of the detected intensity and the observed antiviral potency of the complex extract was calculated for each single m/z feature. We considered correlation coefficients >0.8 as relevant. Principal Component Analysis (PCA) models have been used for data visualization and for discovery of natural occurring patterns as well as for identification of putative outliers. Mean centering in combination with unit variance scaling has been applied for the data of this study (SIMCA-P<sup>®</sup> 9 (Umetrics, Umea, Sweden)). After exploration of naturally occurring patterns with unsupervised methods, the data has been further analyzed with Partial Least Square Discriminant Analysis (PLS-DA)

Panel	Cellular target	Description	Staining solutions
Panel 1	Nucleus	Nuclear parameter	Hoechst33342 (Thermo Fisher Scientific)
	Actin	Cytoskeleton	Phalloidin-FITC (Sigma Aldrich)
	Tubulin	Cytoskeleton	Beta-tubulin antibody and GAM-DyLight 550 (Thermo Fisher Scientific)
	Mitochondria	Mitochondrial membrane potential	MitoTracker <sup>®</sup> Orange CMTMRos (Life Technologies)
Panel 2	Whole cell	Whole cell morphology	Wheat Germ Agglutinin, Alexa Fluor <sup>®</sup> 488 Conjugate (Life Technologies)
	ER	Endoplasmic reticulum	ER-Tracker Blue-White DPX (Life Technologies)
	Lysosomes	Lysosomes	LysoTracker Red DND-99 (Life Technologies)
	Membrane	Membrane	Wheat Germ Agglutinin, Alexa Fluor <sup>®</sup> 488 Conjugate (Life Technologies)
	Nucleus	Nuclear parameters	Hoechst33342 (Thermo Fisher Scientific)
Panel 3	NF-κB	NF-κB activation/distribution	Antibody for NF-κB and GAR-DyLight 550 (Thermo Fisher Scientific)
	Nucleus	Nuclear parameters	Hoechst33342 (Thermo Fisher Scientific)
Panel 4	Caspase 9	Caspase 9 activation/distribution	Caspase 9 antibody and GAM-DyLight 550 (Thermo Fisher Scientific)
	p53	p53 activation/distribution	p53 antibody and GAR- DyLight 488 (Thermo Fisher Scientific)

**Table 2.** Overview of High-content Screening (HCS). Eleven cellular structures were examined by HCS, using 4 panels of of staining solutions with different fluorescent dyes and antibodies.

in SIMCA-P<sup>®</sup> 9 (Umetrics, Umea, Sweden).  $R^2$  was observed to be 0.914. A seven-fold cross-validation has been applied, which resulted in a  $Q^2$  of 0.682. Furthermore, a permutation test for evaluation of putative overfittings has been done with 200 permutations. The regression of the correlation coefficient between the original Y and the permuted Y versus the cumulative  $R^2$  and  $Q^2$  intercepts the x-axis at  $R^2 = 0.0887$  and  $Q^2 = -0.125$ . Therefore, the model has met all the required quality criteria to be considered valid and not overfitting. M/z features with a VIP-VALUE (variable importance in projection) > 1.98 have been considered relevant. Wilcoxon-Mann-Whitney test has been further used to assess the statistical significance of these m/z features (MultiExperimentViewer MeV v4.6.2<sup>59</sup>). P-values have been Benjamini/Hochberg corrected for multi testing. All cytological profiles were normalized to the average of cytological profiles detected for control cells and subjected to hierarchical clustering using complete linkage clustering with optimized gene leaf order and a Pearson correlation using Multi Experiment Viewer (MeV v4.9<sup>59</sup>, Dana-Farber Cancer Institute, Boston, MA, USA).

### Data availability

All data needed to evaluate the conclusions of the paper are present in the paper and/or the Supplementary Material. Additional data related to this paper may be requested from the authors.

Received: 8 February 2019; Accepted: 2 January 2020;

Published online: 28 January 2020

### References

- Gu, J. *et al.* Use of Natural Products as Chemical Library for Drug Discovery and Network Pharmacology. *PLOS ONE* **8**, e62839, <https://doi.org/10.1371/journal.pone.0062839> (2013).
- Newman, D. J., Cragg, G. M. & Snader, K. M. Natural products as sources of new drugs over the period 1981–2002. *J. Nat. Prod.* **66**, 1022–1037, <https://doi.org/10.1021/np030096l> (2003).
- Lam, K. S. New aspects of natural products in drug discovery. *Trends Microbiol* **15**, 279–289, <https://doi.org/10.1016/j.tim.2007.04.001> (2007).
- Newman, D. J. & Cragg, G. M. Natural Products As Sources of New Drugs over the 30 Years from 1981 to 2010. *Journal of Natural Products* **75**, 311–335, <https://doi.org/10.1021/np200906s> (2012).
- Pye, C. R., Bertin, M. J., Lokey, R. S., Gerwick, W. H. & Linington, R. G. Retrospective analysis of natural products provides insights for future discovery trends. *Proc Natl Acad Sci USA* **114**, 5601–5606, <https://doi.org/10.1073/pnas.1614680114> (2017).
- Ju, K.-S. *et al.* Discovery of phosphonic acid natural products by mining the genomes of 10,000 actinomycetes. *Proceedings of the National Academy of Sciences* **112**, 12175–12180, <https://doi.org/10.1073/pnas.1500873112> (2015).
- Walsh, C. T. A chemocentric view of the natural product inventory. *Nature Chemical Biology* **11**, 620–624, <https://doi.org/10.1038/nchembio.1894> (2015).
- Charlop-Powers, Z. *et al.* Global biogeographic sampling of bacterial secondary metabolism. *eLife* **4**, e05048, <https://doi.org/10.7554/eLife.05048> (2015).
- Ochsenkühn, M. A., Schmitt-Kopplin, P., Harir, M. & Amin, S. A. Coral metabolite gradients affect microbial community structures and act as a disease cue. *Communications Biology* **1**, 184, <https://doi.org/10.1038/s42003-018-0189-1> (2018).
- Mopper, K., Stubbins, A., Ritchie, J. D., Bialk, H. M. & Hatcher, P. G. Advanced instrumental approaches for characterization of marine dissolved organic matter: extraction techniques, mass spectrometry, and nuclear magnetic resonance spectroscopy. *Chem Rev* **107**, 419–442, <https://doi.org/10.1021/cr050359b> (2007).
- Stenson, A. C., Landing, W. M., Marshall, A. G. & Cooper, W. T. Ionization and Fragmentation of Humic Substances in Electrospray Ionization Fourier Transform-Ion Cyclotron Resonance Mass Spectrometry. *Analytical Chemistry* **74**, 4397–4409, <https://doi.org/10.1021/ac020019f> (2002).
- Hertkorn, N. *et al.* Natural organic matter and the event horizon of mass spectrometry. *Anal Chem* **80**, 8908–8919, <https://doi.org/10.1021/ac800464g> (2008).
- Swinney, D. C. & Anthony, J. How were new medicines discovered? *Nat Rev Drug Discov* **10**, 507–519, <https://doi.org/10.1038/nrd3480> (2011).
- Bray, M. A. *et al.* Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. *Nat Protoc* **11**, 1757–1774, <https://doi.org/10.1038/nprot.2016.105> (2016).

15. Kremb, S. & Voolstra, C. R. High-resolution phenotypic profiling of natural products-induced effects on the single-cell level. *Scientific reports* **7**, 44472, <https://doi.org/10.1038/srep44472> (2017).
16. Young, D. W. *et al.* Integrating high-content screening and ligand-target prediction to identify mechanism of action. *Nat Chem Biol* **4**, 59–68, <https://doi.org/10.1038/nchembio.2007.53> (2008).
17. Schulze, C. J. *et al.* “Function-first” lead discovery: mode of action profiling of natural product libraries using image-based screening. *Chem Biol* **20**, 285–295, <https://doi.org/10.1016/j.chembiol.2012.12.007> (2013).
18. Kang, J. *et al.* Improving drug discovery with high-content phenotypic screens by systematic selection of reporter cell lines. *Nat Biotechnol* **34**, 70–77, <https://doi.org/10.1038/nbt.3419> (2016).
19. Sutherland, J. J. *et al.* A robust high-content imaging approach for probing the mechanism of action and phenotypic outcomes of cell-cycle modulators. *Mol Cancer Ther* **10**, 242–254, <https://doi.org/10.1158/1535-7163.MCT-10-0720> (2011).
20. Kurita, K. L., Glassey, E. & Linington, R. G. Integration of high-content screening and untargeted metabolomics for comprehensive functional annotation of natural product libraries. *Proc Natl Acad Sci USA* **112**, 11999–12004, <https://doi.org/10.1073/pnas.1507743112> (2015).
21. Kremb, S., Muller, C., Schmitt-Kopplin, P. & Voolstra, C. R. Bioactive Potential of Marine Macroalgae from the Central Red Sea (Saudi Arabia) Assessed by High-Throughput Imaging-Based Phenotypic Profiling. *Marine drugs* **15**, <https://doi.org/10.3390/md15030080> (2017).
22. Helfer, M. *et al.* The root extract of the medicinal plant *Pelargonium sidoides* is a potent HIV-1 attachment inhibitor. *PLoS One* **9**, e87487, <https://doi.org/10.1371/journal.pone.0087487> (2014).
23. O'Rourke, A. *et al.* Alkaloids from the Sponge *Stylissa carteri* Present Prospective Scaffolds for the Inhibition of Human Immunodeficiency Virus 1 (HIV-1). *Marine drugs* **14**, <https://doi.org/10.3390/md14020028> (2016).
24. Rebusburg, S. *et al.* Potent *in vitro* antiviral activity of *Cistus incanus* extract against HIV and Filoviruses targets viral envelope proteins. *Scientific reports* **6**, 20394, <https://doi.org/10.1038/srep20394> (2016).
25. Zhernov, Y. V. *et al.* Supramolecular combinations of humic polyanions as potent microbicides with polymodal anti-HIV-activities. *New Journal of Chemistry* **41**, 212–224, <https://doi.org/10.1039/C6NJ00960C> (2017).
26. Kremb, S. *et al.* EASY-HIT: HIV full-replication technology for broad discovery of multiple classes of HIV inhibitors. *Antimicrob Agents Chemother* **54**, 5257–5268, <https://doi.org/10.1128/AAC.00515-10> (2010).
27. Wishart, D. S. *et al.* HMDB: the Human Metabolome Database. *Nucleic Acids Res* **35**, D521–526, <https://doi.org/10.1093/nar/gkl923> (2007).
28. Wishart, D. S. *et al.* HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res* **46**, D608–D617, <https://doi.org/10.1093/nar/gkx1089> (2018).
29. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27–30, <https://doi.org/10.1093/nar/28.1.27> (2000).
30. Sud, M. *et al.* LMSD: LIPID MAPS structure database. *Nucleic Acids Res* **35**, D527–532, <https://doi.org/10.1093/nar/gkl838> (2007).
31. Witt, M., Fuchser, J. & Koch, B. P. Fragmentation studies of fulvic acids using collision induced dissociation fourier transform ion cyclotron resonance mass spectrometry. *Anal Chem* **81**, 2688–2694, <https://doi.org/10.1021/ac802624s> (2009).
32. Pennisi, E. IMAGING. ‘Cell painting’ highlights responses to drugs and toxins. *Science* **352**, 877–878, <https://doi.org/10.1126/science.352.6288.877> (2016).
33. Bray, M. A. *et al.* A dataset of images and morphological profiles of 30 000 small-molecule treatments using the Cell Painting assay. *Gigascience* **6**, 1–5, <https://doi.org/10.1093/gigascience/giw014> (2017).
34. Nguyen-Hai, N. Naturally Occurring NF- $\kappa$ B Inhibitors. *Mini-Reviews in Medicinal Chemistry* **6**, 945–951, <https://doi.org/10.2174/138955706777934937> (2006).
35. Hiscott, J., Kwon, H. & Genin, P. Hostile takeovers: viral appropriation of the NF-kappaB pathway. *J Clin Invest* **107**, 143–151, <https://doi.org/10.1172/JCI11918> (2001).
36. Kremb, S. *et al.* Aqueous extracts of the marine brown alga *Lobophora variegata* inhibit HIV-1 infection at the level of virus entry into cells. *PLoS One* **9**, e103895, <https://doi.org/10.1371/journal.pone.0103895> (2014).
37. Gonsior, M. E. A. Spectral characterization of chromophoric dissolved organic matter (CDOM) in a fjord (Doubtful Sound, New Zealand). *Aquat. Sci.* **70**, 397–409, <https://doi.org/10.1007/s00027-008-8067-4> (2008).
38. M.S., M., P., D. R. & B., H. P. Endemism, species selection and the origin and distribution of the vascular plant flora of New Zealand. *Journal of Biogeography* **28**, 199–216, <https://doi.org/10.1046/j.1365-2699.2001.00525.x> (2001).
39. Rodrigues, T., Reker, D., Schneider, P. & Schneider, G. Counting on natural products for drug design. *Nat Chem* **8**, 531–541, <https://doi.org/10.1038/nchem.2479> (2016).
40. Rohban, M. H. *et al.* Systematic morphological profiling of human gene and allele function via Cell Painting. *eLife* **6**, <https://doi.org/10.7554/eLife.24060> (2017).
41. Fredericksen, B. L., Wei, B. L., Yao, J., Luo, T. & Garcia, J. V. Inhibition of endosomal/lysosomal degradation increases the infectivity of human immunodeficiency virus. *J Virol* **76**, 11440–11446 (2002).
42. Kim, H. S., Quon, M. J. & Kim, J. A. New insights into the mechanisms of polyphenols beyond antioxidant properties; lessons from the green tea polyphenol, epigallocatechin 3-gallate. *Redox Biol* **2**, 187–195, <https://doi.org/10.1016/j.redox.2013.12.022> (2014).
43. Schmitt, P. *et al.* Simultaneous Determination of Ionization Constants and Isoelectric Points of 12 Hydroxy-s-Triazines by Capillary Zone Electrophoresis and Capillary Isoelectric Focusing. *Anal Chem* **69**, 2559–2566, <https://doi.org/10.1021/ac960945s> (1997).
44. Woods, G. C., Simpson, M. J., Koerner, P. J., Napoli, A. & Simpson, A. J. HILIC-NMR: toward the identification of individual molecular components in dissolved organic matter. *Environ Sci Technol* **45**, 3880–3886, <https://doi.org/10.1021/es103425s> (2011).
45. Schmitt, P., Ketrup, A., Freitag, D. & Garrison, A. W. Flocculation of humic substances with metal ions as followed by capillary zone electrophoresis. *Anal Bioanal Chem* **354**, 915–920, <https://doi.org/10.1007/s0021663540915> (1996).
46. Woods, G. C. *et al.* Online high-performance size exclusion chromatography-nuclear magnetic resonance for the characterization of dissolved organic matter. *Environ Sci Technol* **44**, 624–630, <https://doi.org/10.1021/es903042s> (2010).
47. Perminova, I. V. *et al.* Molecular mapping of sorbent selectivities with respect to isolation of Arctic dissolved organic matter as measured by Fourier transform mass spectrometry. *Environ Sci Technol* **48**, 7461–7468, <https://doi.org/10.1021/es5015423> (2014).
48. Yamashita, Y., McCallister, S. L., Koch, B. P., Gonsior, M. & Jaffé, R. Dynamics of dissolved organic matter in fjord ecosystems: Contributions of terrestrial dissolved organic matter in the deep layer. *Estuarine, Coastal and Shelf Science* **159**, 37–49, <https://doi.org/10.1016/j.ecss.2015.03.024> (2015).
49. Schmitt-Kopplin, P. *et al.* Dissolved organic matter in sea spray: a transfer study from marine surface water to aerosols. *Biogeosciences* **9**, 1571–1582, <https://doi.org/10.5194/bg-9-1571-2012> (2012).
50. Ksionzek, K. B. *et al.* Dissolved organic sulfur in the ocean: Biogeochemistry of a petagram inventory. *Science* **354**, 456–459, <https://doi.org/10.1126/science.aaf7796> (2016).
51. Gonsior, M. *et al.* Molecular characterization of effluent organic matter identified by ultrahigh resolution mass spectrometry. *Water Res* **45**, 2943–2953, <https://doi.org/10.1016/j.watres.2011.03.016> (2011).
52. Dittmar, T., Koch, B., Hertkorn, N. & Kattner, G. A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. *Limnology and Oceanography: Methods* **6**, 230–235, <https://doi.org/10.4319/lom.2008.6.230> (2008).
53. Li, Y. *et al.* How representative are dissolved organic matter (DOM) extracts? A comprehensive study of sorbent selectivity for DOM isolation. *Water Res* **116**, 316–323, <https://doi.org/10.1016/j.watres.2017.03.038> (2017).

54. Martínez-Pérez, A. M. *et al.* Linking optical and molecular signatures of dissolved organic matter in the Mediterranean Sea. *Scientific reports* **7**, 3436, <https://doi.org/10.1038/s41598-017-03735-4> (2017).
55. Tziotis, D., Hertkorn, N. & Schmitt-Kopplin, P. Kendrick-analogous network visualisation of ion cyclotron resonance Fourier transform mass spectra: improved options for the assignment of elemental compositions and the classification of organic molecular complexity. *Eur J Mass Spectrom (Chichester)* **17**, 415–421, <https://doi.org/10.1255/ejms.1135> (2011).
56. Kind, T. & Fiehn, O. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinformatics* **8**, 105, <https://doi.org/10.1186/1471-2105-8-105> (2007).
57. Suhre, K. & Schmitt-Kopplin, P. MassTRIX: mass translator into pathways. *Nucleic Acids Res* **36**, W481–484, <https://doi.org/10.1093/nar/gkn194> (2008).
58. Wagele, B., Witting, M., Schmitt-Kopplin, P. & Suhre, K. MassTRIX reloaded: combined analysis and visualization of transcriptome and metabolome data. *PLoS One* **7**, e39860, <https://doi.org/10.1371/journal.pone.0039860> (2012).
59. Saeed, A. I. *et al.* TM4: A Free, Open-Source System for Microarray Data Management and Analysis. *BioTechniques* **34**, 374–378, <https://doi.org/10.2144/03342mt01> (2003).

### Author contributions

C.M., S.K., and P.S.-K. designed the research, M.G. and P.S.-K. collected and extracted samples, C.M., S.K., and P.S.-K. performed the experiments and analyzed the data. C.M., S.K., P.S.-K., M.G., C.V. and R.B.-W. discussed the data and wrote the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-57709-0>.

**Correspondence** and requests for materials should be addressed to C.M. or P.S.-K.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020